3P

УДК 004.934

В.В. Робейко, М.М. Сажок
Міжнародний науково-навчальний центр інформаційних технологій та систем,
м. Київ, Україна
{valya.robeiko, sazhok}@gmail.com

Багаторівнева багатозначна модель перетворення орфографічного тексту на фонемний

У статті розглядається актуальне питання породження послідовностей фонем за орфографічним текстом і зворотне перетворення. Пропонується модель, у якій закладені закономірності зв'язку між фонетичними й орфографічними символами. Багатозначні перетворення, що здійснюються згідно з моделлю, еквівалентні побудові направленого графа. Параметри моделі задаються таблично у вигляді контекстно-залежних правил, які формує експерт. У моделі передбачено введення додаткових рівнів транскрибування, що дає змогу експертові будувати складніші перетворення, працюючи з відносно простими правилами. Розроблено систему, на вхід якої подається орфографічний текст із позначеннями наголосу, а на виході отримуємо послідовності фонем, які відповідають різним варіантам вимови вхідного тексту. Наводяться приклади практичного застосування системи для автоматичного розпізнавання українського спонтанного мовлення.

Вступ

Породження послідовностей фонем за орфографічним текстом і навпаки (породження послідовностей слів за фонемним текстом) є актуальним питанням у царині усномовної інформатики. Графемно-фонемні перетворення потрібні для транскрибування текстів у системах озвучення інформації, для формування словників вимови під час оцінки параметрів акустичної моделі в різних схемах декодування усномовного сигналу тощо [1-4].

Перетворення орфографічного тексту (графем) на фонемний текст (транскрипцію, зроблену за допомогою фонем) неможливе без вивчення закономірностей зв'язку між фонетичними та орфографічними символами.

Тривалий час графемно-фонемне перетворення реалізовувалось у вигляді певного алгоритму – транскриптора, що відтворює у програмному коді правила читання, узяті з підручника або довідника [5].

Потрібно зазначити, що, на відміну від орфографічного, фонемний текст для кожного конкретного слова не є усталеним як із суб'єктивних, так і з об'єктивних причин.

По-перше, до цього часу серед дослідників у царині фонетики та фонології тривають дискусії щодо алфавіту фонем української мови (незважаючи на прийнятий стандарт Міжнародної фонетичної асоціації – *IPA*). По-друге, алфавіт фонем для системи розпізнавання або синтезу мовлення залежить не лише від кількості фонем мови, а й від того, яке фонетичне явище (асиміляція, редукція, палаталізація та ін.) враховується на рівні символів, а яке – на рівні акустичної моделі шляхом уведення контекстної залежності (*CD*-фонеми, фонеми-трифони), використання суміші гаусоїдів (*GMM*) чи зміни топології акустичної моделі *HMM*. По-третє, під час розробки мовленнєвих технологій повинні враховуватися індивідуальні, ситуативні особливості вимови, вимова слів у потоці мовлення, а це спричиняє введення багатозначності при переході до фонемного тексту.

Відсутність розмежування між оперативним кодом (командами) і даними є серйозною перешкодою для модифікації та вдосконалення транскриптора, тобто перетворення його на гнучкий інструмент дослідника.

Задача зворотного переходу від фонемного до орфографічного тексту є відносно новою, вона набула актуальності з розвитком багатозначної багаторівневої моделі розуміння мовленнєвого сигналу [4]. Реалізованому на той час алгоритму подавалося на вхід менше 30 правил для перетворення «фонема-графема», у результаті чого отримувалися гіпотетичні варіанти орфографічного тексту для багатозначної відповіді розпізнавання вільного порядку слідування фонем. Така незначна кількість правил пояснюється властивостями української орфографії, адже правила написання для української мови значною мірою засновуються на фонетичному принципі (пишемо те, що чуємо).

Подальші дослідження алгоритму полягали в його використанні у зворотному процесі – перетворенні орфографічного тексту на фонемний. Мета таких досліджень – отримати транскриптор, який можна порівняно легко модифікувати на рівні структур даних без нової компіляції програмного коду.

Зауважимо, що у світі для перетворення графем на фонеми використовуються статистичні методи, які аналізують створений експертами словник вимови [6]. Це виправдано для мов, у яких орфографічне написання слів суттєво відрізняється від їх вимови (наприклад, для англійської чи французької мови).

У розділі 2 пропонується спосіб опису закономірностей, що пов'язують орфографічні і фонемні символи, та аналізується побудова графа, на якому породжуються гіпотетичні перетворення послідовностей символів. Розділ 3 присвячено комп'ютерній моделі алгоритму та практичному застосуванню системи транскрибування. У висновках обговорюються результати та подальші дослідження.

Модель багатозначного перетворення послідовностей символів

Нехай маємо скінченну послідовність символів

$$\left(a_1, a_2, ..., a_n, ..., a_N\right) \equiv a_1^N,\ a_n \in \mathbf{A}, \qquad (1)$$

де $\mathbf{A}$ – алфавіт вхідних символів. Сконструюємо відображення цієї послідовності на множину послідовностей вихідних символів із деякого іншого алфавіту $\mathbf{B}$.

Розглянемо функцію f, що відображає послідовність a_1^N, починаючи з її n-го символу, у символ алфавіту $\mathbf{B}$ або порожню множину:

$$f : a_n^N \to b,\ \ b \in \mathbf{B} \cup \varnothing,\ \ 1 \le n \le N. \qquad (2)$$

Зауважимо, що (2) має місце лише у випадку, коли вхідна послідовність належить області визначення f, тобто $a_n^N \in \mathrm{Def}\left(f\right)$. Множина послідовних застосувань таких функцій переводить a_n^N у послідовності символів з алфавіту $\mathbf{B}$, утворюючи таким чином мультифункцію:

$$\mathrm{F}\left(a_n^N\right) = \left\{\left(f_1^k\left(a_n^N\right), f_2^k\left(a_n^N\right), ..., f_{L_k}^k\left(a_n^N\right)\right) \in \mathbf{B}^{L_k} \cup \varnothing,\ 1 \le k \le K_{\mathrm{F}}\right\}, \qquad (3)$$

де L_k – довжина k-ї вихідної послідовності, загальна кількість яких, K_{F}, своя для кожної $\mathrm{F} \in \mathbf{F}$.

Визначимо аналог прямого добутку над множинами, отриманими внаслідок дії мультифункцій з $\mathbf{F}$, як перебір усіх варіантів об'єднання скінченних послідовностей символів з алфавіту $\mathbf{B}$. Тобто, опускаючи аргументи мультифункцій:

$$\mathrm{F} \otimes \mathrm{G} = \left\{\left(f_1^u, f_2^u, ..., f_{L_u}^u, g_1^v, g_2^v, ..., g_{L_v}^v\right),\ 1 \le u \le K_{\mathrm{F}},\ 1 \le v \le K_{\mathrm{G}}\right\}. \qquad (4)$$

Припускаємо за визначенням, що якщо результат дії F або G є порожньою множиною, то результатом їх добутку буде порожня множина. На відміну від декартового добутку для визначеного нами аналога виконується властивість асоціативності.

Розглянемо впорядковану множину $\widetilde{\mathbf{F}}$ мультифункцій $\mathrm{F} \in \mathbf{F}$, які супроводимо додатковими параметрами:

$$\widetilde{\mathbf{F}} = \left(\mathrm{F}_{i,d_i,\delta_i}\right),\ 1 \le i \le \left|\widetilde{\mathbf{F}}\right|,\ d_i > 0,\ \delta_i = \{0,1\}\,, \tag{5}$$

де i є індексом мультифункції у впорядкованій множині $\widetilde{\mathbf{F}}$; параметр d_i назвемо шириною кроку аналізу, δ_i – «умовою виключності». Через ці параметри конятруюємо обмеження при обчисленні добутку

$$\bigotimes_{i,n} \mathrm{F}_{i,d_i,\delta_i}\left(a_n^N\right),\ 1 \le i \le \left|\widetilde{\mathbf{F}}\right|,\ 1 \le n \le N\,. \tag{6}$$

Припустимо, що ми вже обчислили вираз (6) на деяких упорядкованих індексних множинах J і M і отримали деяку непорожню множину

$$G_{J,M} = \bigotimes_{u \in J, v \in M} \mathrm{F}_{u,d_u,\delta_u}\left(a_v^N\right)\,. \tag{7}$$

Нехай j та m є останніми елементами індексних множин J і M відповідно. Тоді при розгляді наступної компоненти добутку, $\mathrm{F}_{i,d_i,\delta_i}\left(a_n^N\right)$, проводимо обчислення згідно з визначенням (4), якщо виконуються такі умови:

$$\begin{cases} m + d_j = n; \\ \left[\begin{array}{l} \delta_r \ne 1,\ 1 \le r < i; \\ \bigotimes\limits_{u \in J, v \in M} \mathrm{F}_{u,d_u,\delta_u}\left(a_v^N\right) \otimes \mathrm{F}_{r,d_r,\delta_r}\left(a_n^N\right) \ne \varnothing,\ 1 \le r < i,\ \text{якщо}\ \delta_i = 1. \end{array}\right. \end{cases} \tag{8}$$

В іншому випадку, при надходженні наступної компоненти добутку отримуємо порожню множину.

Виразом (6) породжуються послідовності вихідних символів за деякою послідовністю вхідних символів. Якщо вхідний алфавіт збігається з алфавітом літер певної мови, а вихідний алфавіт складається з фонем, то маємо багатозначний транскриптор орфографічного тексту. І навпаки, якщо на вході – фонемний алфавіт, а на виході – алфавіт літер, то отримаємо багатозначне перетворення з фонемного тексту на орфографічний. Можливі проміжні варіанти.

Приклад породження варіантів фонемного тексту за орфографічним наведено на рис. 1. Розглядається орфографічний текст одного слова «сніг». Із метою узагальнення стандартний алфавіт української мови (включно з апострофом) доповнено символом “_”, який розділяє слова. Літери, що відповідають наголошеним голосним, переводяться у верхній регістр, всі інші літери – у нижній. Позиція наголосу визначається за орфоепічним словником з урахуванням омографії [7] або беруться до уваги всі допустимі позиції наголосу.

Отже маємо на вході послідовність із шести символів a_1^N = (“_”, “*с*”, “*н*”, “*І*”, “*г*”, “_”), $N = 6$. На графі відображено всі допустимі мультифункції $\mathrm{F}_{i,d_i,\delta_i}\left(a_n^N\right), 1 \le n \le N$. Здійснюючи перехід за стрілками, отримуємо добутки вигляду (6), які генерують чотири послідовності фонем або фонемні тексти:

“_ *с н' І х* _”; “_ *с н' І г* _”; “_ *с' н' І х* _”; “_ *с' н' І г* _”.

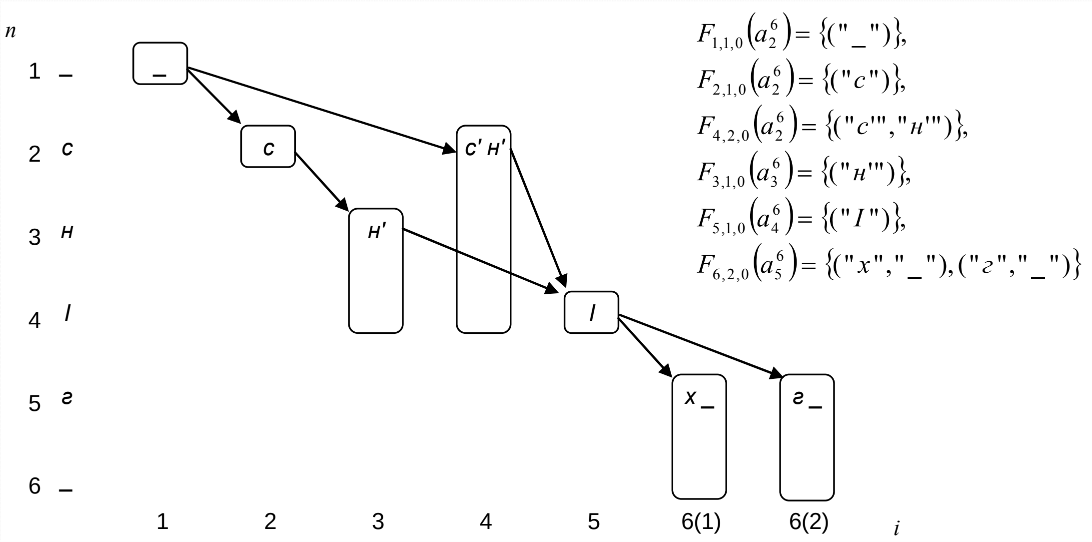


Рисунок 1 – Граф багатозначного перетворення
орфографічного слова «сніг» на фонемну транскрипцію

Зауважимо, що на практиці використовується не вся підпослідовність вигляду a_n^N, а лише a_n^{n-1+T}, де додатній параметр T залежить від конкретної мультифункції і визначає ширину контексту, у якому розглядаються вхідні символи. На рис. 1 висота прямокутників у вузлах графа відповідає ширині контексту.

Результат перетворення може бути багаторазово підданий описаним перетворенням з тими самими або іншими параметрами. Таким чином ми можемо вводити додаткові рівні транскрибування, на яких відбувається спрощення специфікації параметрів, що важливо для мов, у яких написання відчутно відрізняється від вимови.

Система багаторівневого багатозначного транскрибування орфографічних текстів

Пропонується система, на вхід якої подається орфографічний текст, що містить лише символи з алфавіту літер включно з символами границі між словами та морфемами та позначеннями наголосу. Такий текст отримуємо внаслідок автоматичного оброблення довільного тексту або списку слів, що полягає у вербалізації (розшифруванні) символів, чисел і скорочень, розставлянні наголосів і поділу на синтагми, слова та морфеми. На виході системи отримуємо послідовності фонем, що відповідають різним варіантам вимови вхідного тексту. Ця система транскрибування використовується для розпізнавання українського мовлення [8], [9].

Розроблена система багатозначного транскрибування орфографічних текстів використовує модель (1) – (8), що дає змогу таблично задавати контекстно-залежні правила перетворень одних узагальнених послідовностей символів на інші. При цьому в кожному правилі задається ширина кроку, за яким відбувається перехід до наступної послідовності символів. Також є можливим застосування багатьох правил, що дає змогу генерувати одразу декілька варіантів транскрипції одного й того самого слова (або генерувати потрібний варіант із кількох можливих, наприклад, описуючи спонтанне мовлення мовця чи групи мовців).

Наведемо деякі правила з поясненнями у табл. 1.

Таблиця 1 – Приклади побудови правил транскриптора

Вхідна підпослідовність символів	Вихідна послідовність символів	Ширина кроку	Пояснення
[зсц] [жшч]	[жшч]	1	**з**, **с**, **ц** перед **ж**, **ш**, **ч** переходять відповідно у **ж**, **ш**, **ч**
т[дтзснц][іІєюяЄЮЯь]	т'	1	**т** перед м’якими **д**, **т**, **з**, **с**, **н**, **ц** пом’якшується
с т [лн]	с	2	**т** між **с** та **л** або **н** випадає

Загалом для транскрибування (для літературного варіанта вимови) достатньо ввести близько 30 подібних правил.

Тестування системи транскрибування відбувалося на основі розпізнавання спонтанного мовлення кількох сотень дикторів [8], [9].

Для всіх дикторів був створений загальний варіант транскрибування на основі літературної вимови. Окрім цього, всі диктори були розподілені на групи, для яких розроблено свої правила індивідуалізованого транскрибування, що доповнюють або заміняють основний варіант [8]. Також для однієї із груп дикторів був розроблений загальний варіант транскрибування спонтанного злитого мовлення [9].

Результати вивчення спонтанного мовлення багатьох дикторів свідчать про те, що ніхто з них не дотримується орфоепічних правил вимови у повному обсязі. Насамперед це стосується заборонених літературною нормою регресивної асиміляції за глухістю в парі фонем «дзвінка+глуха» й оглушення приголосних перед паузою (**тобто → т о п т о**; **підтримати → п' і т т р И м а т и**; **робив → р о- б И ф**). Диктори з такими особливостями вимови були виділені в окрему групу. Було виокремлено й багато інших характерних рис вимови різних дикторів: редукція закінчень деяких слів (прикметників, дієслів) у злитому мовленні (**шановний → ш а н О в н и; доброго → д О б р о**), «акання» (**робити → р а б И т и**), тверда вимова м’яких приголосних (**синього → с И н о г о**) та ін.

Для деяких слів (службових частин мови, слів із різними наголосами, наприклад) задається декілька варіантів транскрипцій – із наголосом на різних складах (якщо в мові можливі різні варіанти прочитання таких слів) або взагалі без наголосу: **коли → к о л И ; к О л и ; к о л и**.

Такі тенденції моделюються шляхом зміни правил переходу від одних послідовностей символів до інших і розширенням чинних правил.

Усі правила індивідуалізованої модифікації транскрипцій можна розділити на декілька груп (на основі опрацювання звукових записів).

До позиційних (змін, що залежать від загальних фонетичних умов – позиції у складі/слові, наголошеності/ненаголошеності тощо [10]) нелітературних змін звуків у потоці мовлення зараховуємо:

1) ослаблену вимову *о* як *а* в ненаголошеній позиції, рідше трапляється редукція ненаголошених голосних до повного зникнення (**тепер → т и п Е р, зозуля → з у з У л' а, боротьба → б а р а д' б А** або **б р а д' б А**);

2) оглушення дзвінких приголосних перед паузою (**брід → б р' І т, зараз → з А р а с**);

3) редукцію у термінальних частинах слів у процесі мовлення – зникнення приголосного звука в закінченнях *-ого*, *-их*, *-ич*, *-ий*, *-іх*, *-ій*, *-ії*, *-ої*, *-еї*, *-ою*, *-єю*, *-ити* та

подібних (**коротший → к о р О ч ш и, Іванович → і в А н о в и, синіх → с И- н' і, безпекою → б е с п Е к о у**); зникнення кінцевого голосного звука в закінченнях *-ою*, *-ею*, *-єю* та подібних (**доброю → д О б р о й, землею → з е м л Е й**) та ін.

До комбінаторних (якісні та кількісні зміни сусідніх звуків [10]) нелітературних змін відносимо:

1) повну регресивну асиміляцію за глухістю у сполуці «дзвінкий+глухий» на межі будь-яких морфем у слові та на межі слів (**без причини → б е с п р и ч И н и, розсунути → р о с с У н у т и , книжка → к н И ш к а, сядьте → с' А т' т е**);

2) асиміляцію за м'якістю свистячих та шиплячих приголосних, губних та задньоязикових приголосних (**злі → з' л' І, шлях → ш' л' А х, квітка → к' в' І т к а**);

3) вимову подовжених приголосних звуків як звичайного неподовженого звука, вимову двох голосних як одного звука (**віддати → в' і д А т и, знання → з на- н' А, зоопарк → з о п А р к, аеропорт → а р о п О р т**);

4) неповне спрощення в групах приголосних, його відсутність (**чесний → ч Е с т н и й**) та ін.

Для створення індивідуалізованих словників вимови всі досліджувані диктори були розподілені на групи, що відбувалося в кілька етапів: перший передбачав виокремлення групи дикторів із літературним мовленням та дикторів із відхиленнями від літературної норми; наступною ділянкою роботи було виділення найзагальніших спільних рис мовлення у другій групі та створення відповідних транскрипцій для кожної з підгруп. Таким чином було згенеровано 18 видів транскрипцій словника для тих чи інших дикторів. Адекватність індивідуалізованих транскрипцій перевірялася за допомогою розпізнавання мовлення (покращення/погіршення результатів розпізнавання для конкретного диктора у порівнянні з результатами розпізнавання для цього ж диктора з використанням літературної транскрипції). Кожна з транскрипцій була перевірена на всіх дикторах. Приклади словників – у табл. 2.

Таблиця 2 – Приклади індивідуалізованих словників спонтанного мовлення

Індивідуалізований словник / Літературна транскрипція	devocal	end_cons	a	devocal _ a
р о б И в	р о б И ф	р о б И	р а б И в	р а б И ф
в О р о г	в О р о х	в О р о	в О р а г	в О р а х

Для досліджень вимови (і для навчальної вибірки під час розпізнавання) використовувалися записи засідань Верховної Ради України тривалістю 99 тис. с, у яких зустрілося понад 210 тис. слів. Усього було записано 208 дикторів. Дикторів із тривалістю запису понад 300 с виявилося 87 (запис тривалістю менше 300 с є занадто коротким для точних висновків у процесі розпізнавання) [8].

Для перевірки правильності транскрипцій (і для контрольної вибірки під час розпізнавання) використовувалися записи тривалістю 30 тис. с, у яких зустрілося понад 68 тис. слів. Ці записи були зроблені у відмінні від навчальної вибірки дні. Всього використовувалися записи 118 дикторів. Дикторів із тривалістю запису понад 300 с виявилося 37.

Словник для генерування транскрипцій був створений із текстів стенограм засідань Верховної Ради України. Із офіційного сайту Верховної Ради були завантажені стенограми засідань, починаючи з 1991 р. Усі тексти стенограм (понад 15 млн слів) були модифіковані для того, щоб усунути зайву інформацію (наприклад, повідомлення про

аплодисменти чи вигуки), записати числа словами, а також відокремити український текст від російського. У результаті було отримано словник із приблизно 150 тис. слів, для якого обчислено частоту вживаності кожного слова. Було укладено словник на 15 тис. одиниць із найчастотніших слів (50 і більше слововживань), який і став основою для аналізу.

Щоб з'ясувати, наскільки впливає індивідуалізація транскрипції на надійність розпізнавання, було проведено ряд експериментів. Для цього зроблено оцінку параметрів акустичної та лінгвістичної моделей [8] із використанням інструментарію [2]. Результати експериментів розпізнавання для деяких транскрипцій (покращення/погіршення результатів розпізнавання для конкретного диктора у порівнянні з результатами розпізнавання для цього ж диктора з використанням літературної транскрипції) наведені у табл. 3.

Таблиця 3 – Зразки результатів розпізнавання мовлення дикторів із використанням індивідуалізованих транскрипцій

Словник / Диктор	end_vow_rus	end_vow	end_cons_2	end_cons_1	a	duples
lit	0,16	0,14	-0,01	-0,01	-2,6	0,26
kir	1,37	1,82	0	-1,06	-3,3	0,91
dan	0,71	1,16	-0,09	-0,44	-5,4	0,18

Після опрацювання результатів експериментів кожному з дикторів приписуються ті правила транскрибування мовлення, які підвищують надійність розпізнавання. Таким чином формується набір правил для генерування індивідуалізованих транскрипцій того чи іншого мовця.

Як виявилося, частина правил транскрибування може бути використана для більшості мовців, наприклад, редукція закінчень унаслідок швидкого темпу та емоційності мовлення. Дуже поширеним явищем є оглушення, спричинене впливом російської мови. Значна кількість дикторів уникає вимови подовжених та подвоєних звуків. Часто зустрічається слабка вимова наголошених голосних (наголос зникає не тільки в односкладових чи двоскладових словах, а й у трискладових). Інші правила можуть бути використані лише для мовлення одного-двох дикторів.

Генерування індивідуалізованих транскрипцій для розпізнавання мовлення дозволило покращити його надійність. Надалі планується збільшити кількість індивідуалізованих транскрипцій за рахунок детальнішого аналізу мовлення дикторів, а також автоматично заносити мовців до тієї чи іншої групи, урахувавши всі особливості їх вимови.

Загальний словник транскрипцій для спонтанного мовлення був створений для розпізнавання мовлення дикторів акустичного корпусу українського ефірного мовлення (АКУЕМ) [11]. Він поєднує у собі як правила літературної вимови, так і індивідуальні риси вимови окремих дикторів (враховуючи можливість неоднозначної вимови диктором того самого слова), а також окреслює специфіку спонтанного українського мовлення (наприклад, задає наголошену та ненаголошену форми для всіх односкладових слів тощо).

Для досліджень вимови (і для навчальної вибірки під час розпізнавання) використовувалися звукові записи спонтанного мовлення із АКУЕМ тривалістю біля 190 тис. с (понад 2 000 дикторів). Особлива увага зверталася на мовлення дикторів із тривалістю запису понад 600 с (близько 150 дикторів).

Словник для генерування транскрипцій був створений із текстів стенограм АКУЕМ – близько 50 тис. слів.

Для перевірки правильності транскрипцій (і для контрольної вибірки під час розпізнавання) використовувалися записи тривалістю 13,5 тис. с, у яких зустрілося 29 500 слів. Всього використовувалися записи 34 дикторів із середнім та швидким темпом мовлення.

Можливість генерувати одразу декілька варіантів транскрипції одного й того самого слова дозволяє продемонструвати в словнику варіативність вимови найчастотніших українських слів, редукування та розтягування слів під час швидкого темпу мовлення, нечітку вимову та подібні явища поряд із літературним варіантом вимови. Також система транскрибування дає змогу генерувати транскрипції для таких специфічних підсловників, як словник суржику, соціальних та територіальних діалектів, абревіатур та ін.

Використання варіативного словника для розпізнавання спонтанного мовлення покращує надійність його розпізнавання [9].

Надалі планується збалансувати набір правил, які породжують варіанти вимови слів у потоці спонтанного мовлення, а також використати індивідуалізовані словники (для 18 груп дикторів) під час створення єдиного словника для розпізнавання спонтанного українського мовлення.

Висновки

Запропонована модель дає змогу досить компактно описати закономірності перетворень між графемами та фонемами у зручному для експерта вигляді з урахуванням неоднозначності перетворень. Розроблена реалізація моделі придатна також для розшифрування чисел, символів і скорочень.

Уведення у модель довільної кількості рівнів транскрибування допомагає розкласти правила, які задає експерт, на простіші. Це відкриває шлях до застосування запропонованого методу для транскрибування мов, у яких традиції *написання* не відповідають вимові слів. У спонтанному українському мовленні спостерігається еквівалентне явище, але у зворотному напрямку: відхід реальної живої *вимови* від написання (і, відповідно, від правил канонічної літературної вимови). У системі розпізнавання спонтанного мовлення потрібно знайти баланс між детальністю фонемної транскрипції, топологією акустичних моделей фонем та кількістю параметрів, які уточнюють ці моделі.

Результатом багатозначного розв'язку є послідовності символів, але при цьому не вказується, які з них більш імовірні. Оцінка відповідних імовірностей можлива на основі результатів розпізнавання, що є предметом подальших досліджень у цій сфері.

Література

1. Винцюк Т.К. Анализ, распознавание и смысловая интерпретация речевых сигналов / Винцюк Т.К. – Киев : Наукова думка, 1987. – 264 с.
2. HTK Book, version 3.1 / [Young S.J. et al.]. – Cambridge University, 2002.
3. Lee. Julius – an open source real-time large vocabulary recognition engine / Lee, T. Kawahara, K. Shikano // In Proc. European Conference on Speech Communication and Technology (EUROSPEECH), 2001. – P. 1691-1694.
4. Taras Vintsiuk. Multi-Level Multi-Decision Models for ASR / Taras Vintsiuk Mykola Sazhok // Proceedings of the 10th Int. Conference on Speech and Computer – SpeCom'2005, Patras, 2005. –P. 69-76.
5. Автоматичний озвучувач українських текстів на основі фонемно-трифонної моделі з використанням природного мовного сигналу / [Тарас Вінцюк, Тетяна Людовик, Микола Сажок, Руслан Селюх] // Праці 6-ї Всеукраїнської міжнародної конференції : «Оброблення сигналів і зображень та розпізнавання образів» –Київ : УкрОбраз'2002,2002.
6. Bisani M. Joint-sequence models for grapheme-to-phoneme conversion / M. Bisani, H. Ney // Journal Speech Communication. – Elsevier, 2008. – № 50. – P. 434-451.

7. Робейко В.В. Графічна омонімія як специфічна проблема синтезу мовлення за текстом / В.В. Робейко // Українське мовознавство. – Вип. 39/1. – Київ, 2009. – С. 429-433.
8. Пилипенко В.В. Автоматизированный стенограф украинской речи / В.В. Пилипенко, В.В. Робейко // Штучний інтелект. – № 4. – 2008. – С. 768-775.
9. Людовик Т.В. Автоматическое распознавание спонтанной украинской речи (на материале корпуса украинской эфирной речи) / Т.В. Людовик, В.В. Пилипенко , В.В. Робейко // Компьютерная лингвистика и интеллектуальные технологи : по материалам ежегодной Международной конференции «Диалог». –. М.: РГГУ, 2011. – Вип. 10 (17). – С. 478-488.
10. Сучасна українська літературна мова. Фонетика : [навч. посібник для студентів-філологів]. – К. : Видавничо-поліграфічний центр «Київський університет», 2002. – С. 60.
11. Створення акустичного корпусу українського ефірного мовлення / [Н.Б. Васильєва В.В. Пилипенко, О.М. Радуцький та інш.] // Обробка сигналів і зображень та розпізнавання образів : Десята Всеукраїнська міжнародна конференція. – Київ, 2010. – С. 55-58.
12. Робейко В.В. Генерування індивідуалізованих транскрипцій для злитого мовлення / В.В. Робейко // Мовні і концептуальні картини світу. – Вип. 26, Ч. 3. – Київ, 2009. – С. 38-42.

Literatura

1. Vincju T.K. Analiz, raspoznavanie i smyslovaja interpretacija rechevyh signalov. Kiev: Naukova dumka. 1987. 264 s.
2. Young S.J. HTK Book, version 3.1. Cambridge University, 2002
3. Lee T. Proc. European Conference on Speech Communication and Technology (EUROSPEECH). 2001. P 1691-1694
4. Vintsiuk T. Proceedings of the 10th Int. Conference on Speech and Computer – SpeCom’2005. Patras. 2005. P 69-76
5. Bisani M. Journal Speech Communication № 50. Elsevier. 2008. P 434-451
6. Vincyuk T. Praci 6-yi Vseukrayins’koyi mizhnarodnoyi konferenciyi “Obroblennya syhnaliv i zobrazhen’ ta rozpiznavannya obraziv”. UkrObraz’2002. Kyiv. 2002
7. Suchasna ukrayins’ka literaturna mova. Fonetyka: Navch. posibnyk dlya studentiv-filolohiv. K.: Vydavnycho-polihrafichnyj centr “Kyivs’kyj universytet”. 2002. S 60
8. Pylypenko V.V. Shtuchnyj intelekt. № 4. 2008. S 768-775
9. Robejko V.V. Movni i konceptual’ni kartyny svitu. Vyp. 26. Ch. 3. Kyiv. 2009. S 38-42
10. Robejko V.V. Ukrayins’ke movoznavstvo. Vyp. 39/1. Kyiv. 2009. S 429-433
11. Ljudovik T.V. Komp’juternaja lingvistika i intellektual’nye tehnologii: Po materialam ezhegodnoj Mezhdunarodnoj konferencii “Dialog”. Vyp. 10 (17). M.: RGGU. 2011. S 478-488
12. Vasyl’yeva N.B. Obrobka syhnaliv i zobrazhen’ ta rozpiznavannya obraziv: Desyata Vseukrayins’ka mizhnarodna konferenciya. Kyiv. 2010. S 55-58

В.В. Робейко, М.М. Сажок

Многоуровневая многозначная модель превращения орфографического текста в фонемный

В статье рассматривается актуальный вопрос порождения последовательностей фонем по орфографическому тексту и обратное преобразование. Предлагается модель, в которой заложены закономерности связи между фонетическими и орфографическими символами. Многозначные преобразования, осуществляемые согласно модели, эквивалентны построению направленного графа. Параметры модели задаются таблично в виде контекстно-зависимых правил, которые формирует эксперт. В модели предусмотрено введение дополнительных уровней транскрибирования, что позволяет эксперту строить сложные преобразования, работая с относительно простыми правилами. Разработана система, на вход которой подается орфографический текст с обозначениями ударений, а на выходе получаем последовательности фонем, соответствующих различным вариантам произношения исходного текста. Приводятся примеры практического применения системы для автоматического распознавания украинской спонтанной речи.

V.V. Robejko, M.M. Sazhok

Multilevel Multivalued Model for Grapheme-to-Phoneme Text Conversion

In this paper we consider the actual problem of grapheme-to-phoneme conversion for forward and backward directions. We propose the model describing regularities of relations between the phonetic and spelling symbols. Multi-decision transformations carried out according to the model, are equivalent to building of a directed graph. The model parameters are specified in tabular form as a set of context-sensitive rules formed by the expert. Introduced additional model levels of transcription allow the expert to build complex transformations, still working with relatively simple rules. The developed system takes the orthographic text with the pointed accentuation and produces phoneme sequences corresponding to the different pronunciation variants of the input text. The examples of practical applications of the system for automatic recognition of Ukrainian spontaneous speech are described.

Стаття надійшла до редакції 22.06.2011.